The 2nd Workshop on Deep Models and Artificial Intelligence for Defense Applications: Potentials, Theories, Practices, Tools and Risks
November 11-12, 2020

Summary

By Ying Zhao, Erik Blasch, Doug Lange, Tony Kendall, Arjuna Flenner, Bonnie Johnson, and Bruce Nagy

This workshop addressed two key issues: AI challenges and uniqueness of defense applications.

Challenges: Advancements in hardware, algorithms, and data collection are enabling unexplored defense applications of AI. Development of these applications requires overcoming several challenges. The first challenge is noisy and unstructured data. The second is that adversaries can deceive, corrupt, and camouflage true data; defense applications need to evaluate bad data, find fake data, and perform with limited data. A second challenge is mapping AI algorithms at the strategic, operational, and tactical levels to defense applications [1]. During this mapping, AI applications need to comply with four factors: data; trust; security; and human-machine teaming. In conjunction with AI, data analytics must address the issues of agility, interoperability, and maintainability. Agility of product development includes five topics: open architectures; signal processing; systems software; autonomy via context awareness; and health monitoring. Interoperability is essential for multi-domain coordinated sensing, modeling, and instrumentation. Maintainability enables disaster operations, cyber sensemaking, and predictive maintenance. These topics were discussed through data strategy, algorithms, trust, and standards.

Data strategy: To foster better data collection, the 2020 U.S. DoD data strategy [2] contains seven desirable data elements: visibility; accessible; understandability; linkages; trustworthiness; interoperable; and security. Collected training data must be secured to prevent hostile takeover and made robust against external attacks. Moreover, due to expensive data collections such as battle damage assessment, the DoD needs high-fidelity 3D modeling to generate synthetic training data. The presence of adversaries and unique data requirements necessitates careful consideration of collected and synthetic data.

Algorithms and Technologies: A wide variety of algorithms and their related technologies were discussed. Presenters discussed (co)evolutionary algorithms, game theory, and optimization techniques. Evolutionary algorithms, which do not require gradient computation, can quickly search and evolve to find new battlespace measure/countermeasure configurations and emerging properties. Evolutionary algorithms were also inventively applied to look for tax loop-holes and fixes [7]. Counterfactual regret minimization (CFR) and Alpha-Zero algorithms were highlighted in four applications: AFSIM enabled competitive wargaming simulations; Gomoku; Othello; and DARPA sail-on. Lexical link analysis, an unsupervised learning algorithm, was used to improve prediction and readiness for Navy logistics and supply enterprise. Deep learning was applied to Synthetic Aperture Radar (SAR) images. Interactive machine learning (IML), in a human-machine shared environment, learns human tasks. Lastly, a problem was presented still in need of an algorithmic solution: With implicitly self-similar structures such as fractals, order may emerge from a randomly generated but constrained topology [4].

Many technologies were not tied to any algorithm, such as: cyber malware detection; attack and defense's arms race; multi-segment asymmetrical wargames; strike mission planning; battlespace readiness engagement matrix; and SoarTech's technology for DARPA's AlphaDogFight trials. Two important technologies with needed applications were highlighted: trusted AI and complex system theory. The first technology was used to build warfighter assistants where trusted AI is an automation tool. The second category of complex system theory controlled a swarm in battlefield conditions. This technology was shown to produce millisecond topological pictures of IoT/edge devices over distributed C2/resilient communications within denied environments.

Trust: Mission execution requires trust between the AI enabled and human team members. Due to this importance, many of the talks chose to address trust. The DARPA XAI program discovered that users understand and trust models that match expectations; they even prefer satisfying models over high-performing models. Also, Lipton et al. discussed ten model interpretability dimensions of trust [3]. The diversity of discussions on trust demonstrated that the defense community needs teams including experts on algorithms, design guidance, and best practices to access measures of trust concepts in AI.

Standards: As proposed by the Joint AI center (JAIC), AI systems need standards for responsible, equitable, traceable, reliable, and governable AI systems [5]. A Multisource AI scorecard table (MAST) supports Test and Evaluation, which may be viewed as an initial version of AI application standards. MAST connects governance, explainability and compliance for AI enterprises [6]: Mast adheres to AI defense applications need to be resilient to deception/misclassification, to noisy data, to exploitation of classifiers from known weaknesses and unanticipated attacks.

In conclusion, defense applications tend to be human-in-the-loop, where Defense AI and deep models are a "force multiplier" supporting moral, ethical and legal human decision making.

References:

[1] Lee, K.F. (2018). *AI Superpowers: China, Silicon Valley, and the New World Order*.Publisher*:* Houghton Mifflin Harcourt.

[2] https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF

[3] Lipton, Z.C (2018). The Mythos of Model Interpretability. Presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY. Retrieved from https://arxiv.org/abs/1606.03490

[4] Schaff, J. (2018). Leveraging deterministic chaos to mitigate combinatorial explosions. In *Engineering emergence: a modeling and simulation approach* (Rainey, L. B., Jamshidi, M., eds).CRC press @2019]

[5] https://www.ainfosec.com/wp-content/uploads/2015/09/AIS-SecureView-Overview.pdf

[6] https://arxiv.org/abs/1910.00193

[7]http://dl.acm.org/citation.cfm?doid=2746090.2746099